Response to reviewer 1

We thank reviewer 1 for their comments on our manuscript. The reviewer comments are given below in *italics* with our response in blue. Text additions will be in *italics blue*.

This is a major effort of producing improved global C stocks and fluxes data for a wide range of applications with some success. Yet the new data product have several drawbacks.

We thank the reviewer for recognising that this is a major effort and that it has the potential for a wide range of applications. We address each of the reviewer's concerns below.

Only MODIS products are assimilated, while ignoring ground observations. There are plenty of in-situ observations of fAPAR and GPP available from, e.g. Fluxnet. Satellite retrievals of fAPAR and GPP are indirect with limited resolution, hence subject to algorithm errors and uncertainties. Without ground observations (even at different resolution from those of satellite remote sensing observations), satellite retrievals of GPP, as well as other bio-ecological variables, may have unbounded systematic errors or biases.

We do not assimilate MODIS products alone. We assimilate estimates of woody biomass, soil carbon stocks, leaf mass area, carbon use efficiency and photosynthetic capacity; none of which are MODIS products (Sec. 2.4 and 2.5). As stated in Sec. 2.4 we use LAI, fAPAR and GPP from MODIS to provide consistency in the estimation of these variables – specifically APAR is an input to both LAI and GPP (Line 156-160). However, we do not specifically advocate for the use of MODIS products (Line 160) as e.g. GPP which has been shown to be highly variable between different data products (see Line 44). Each dataset used comes with an estimate of its own uncertainty which we assimilate and are therefore explicitly balanced by our Bayesian analysis.

This study does not directly use in-situ observations as this study is specifically focused on the capacity of current EO datasets to identify sources and sinks in space and time. Each of the EO-derived datasets we assimilate have been extensively calibrated and validated against in-situ networks. Furthermore, the scale mismatch between the analysis (~50x50 km) and of typical in-situ estimate (< 1 x 1 km) will likely introduce a significant representation error leading to a biased comparison – i.e. how representative is any given site of its wider environment. Moreover, GPP from eddy covariance is also not a direct measure, but a partitioning of the net ecosystem exchange of CO_2. The partitioning method used to derive GPP also has potentially significant impact on the resultant GPP estimated (e.g. Stoy et al., 2006).

Multiple studies using CARDAMOM-DALEC have been published utilising site scale data for calibration and/or evaluation. These are contained in the CARDAMOM-DALEC references in the manuscript. However, we recognise that we have not explicitly stated in the introduction the history of CARDAMOM-DALEC.

To address this gap, we will modify the opening to section 2.1 (line 80):

"CARDAMOM is a Bayesian model-data fusion (MDF) framework which has been applied across biomes and from site (Bloom & Williams 2015; Smallman & Williams 2019; Famiglietti et al., 2022), to regional (Exbrayat et al., 2018b; White et al., 2019; Caen et al, 2021; Smallman et al., 2021; Myrgiotis et al., 2022*; Milodowski et al., 2023; Williams et al., 2025) and global scales (Bloom et al., 2016; Friedlingstein et al., 2023). CARDAMOM uses an Adaptive Proposal - Markov Chain Monte Carlo algorithm (AP-MCMC; Haario et al., 2001) to estimate ensembles of parameters for a process-model (DALEC)…."*

Finally, as stated in the Sect. 2.5 several of our observational constraints are derived from in-situ information. The carbon use efficiency is based on a meta-analysis of in-situ estimates to derive a global prior. The prior estimates of leaf mass area and soil carbon stocks are both statistical upscaling of in-situ estimates.

The theoretical foundation of this study is shaky as the data product is based on numerous assumptions. For example, parameter proposals are drawn from uniform prior ranges, CARDAMON analyses conducted for each pixel are repeated three times (chains), realistic approximation of parameter combinations is obtained when three chains are converging, the Gelman-Rubin convergence criterion is used, etc. There is no justification for these assumptions. For example, why are three chains instead of, e.g. five, chains needed? Why is the Gelman-Rubin convergence criterion suitable for this analysis?

Respectfully we disagree that the fundamental assumptions of our analysis are "shaky". We make assumptions but those applied here are justified and widely used and published beyond this group of authors. We will respond to the specific assumptions highlighted by the reviewer below, followed by suggested text alterations where appropriate.

All data assimilation frameworks must make assumptions on the likelihood assigned to any given proposed parameter values. These assumptions influence the outcome of the analysis as they impart an expectation of these values based on prior knowledge and should reflect the reality of the available information (for detailed descriptions of the CARDAMOM approach see Bloom & Williams 2015; Bloom et al., 2016; Worden et al. 2025). Our baseline assumption of sampling from a uniform distribution reflects the poor geospatial information on variables represented by process-model parameters. This lack of knowledge is due to under sampling of variability in space or time and

exacerbated due to difficulty in making many measurements at scale (e.g. rooting depth, tissue turnover rates etc.). Under this assumption we only apply a likelihood cost to proposed parameters which fall outside observable bounds as informed by collaborative databases (e.g. Kattge et al., 2011). However, where we have geospatial information anchored by observations with robust estimates of their uncertainty we can and do assimilate them (e.g. 1-carbon use efficiency = p2, leaf mass area = p17, soil C = p23, Table A1).

The CARDAMOM-DALEC approach conducts pixel-based calibration – an approach described extensively in Bloom et al., (2016) and Worden et al., (2025) among others referenced in the manuscript. The resulting analysis retrieves parameters at pixel scale which reflects the information content of the assimilated observations addressing the geospatial biodiversity challenge outlined in Lines 49-54. The capacity of CARDAMOM to retrieve parameter estimates across large spatial domains and produce geospatial patterns that are ecologically consistent has been demonstrated in many previous studies (e.g. Bloom et al., 2016; Smallman et al., 2021; Williams et al., 2025) including across disturbance gradients (Exbrayat et al., 2018b).

For the other points raised we are using standard assumptions / approaches for Bayesian statistics – outlined in the references used in the manuscript (i.e. Haario 2001; Gelman & Rubin 1992) but also see van de Schoot et al., (2021). For example, checking for convergence between chains is an essential but standard approach in Bayesian statistics. Each chain is initiated with a random starting parameter proposal, and the convergence check ensures that the parameter searches have found a solution that is consistent across chains, and therefore more likely to be representative of a global rather than locally optimal solution. Likewise, the Gelman-Rubin convergence criterion is a long established, widely used (cited >23000 times, Google Scholar accessed 27/08/2026) and standard metric within Bayesian Markov Chain Monte Carlo based approaches. Three chains is the minimum needed to demonstrate that the retrieval is robust (van de Schoot et al., 2021). Running more than three chains offers limited advantage but comes at a considerable computational cost.

The concept of uncertainty propagation is confusing and needs elaboration. Why do uncertainties propagation "from individual pixels to the global scale"? Uncertainties are not some fluids that can flow from one place to another. Instead, uncertainty describes incompleteness of the available information about a variable/parameter. "... we assume uncertainties are fully-correlated when propagating from pixel to global estimates" does not make sense. What does "fully-correlated" mean? Why so?

For a fuller explanation of uncertainty and its propagation in space and time in ecological applications we recommend the Dietze (2017). All observations and models contain uncertainty. When combining uncertain numbers, whether they be in time or space, their respective uncertainties must be propagated. The method for propagating Gaussian uncertainties is well established, though often challenging to implement due to a lack of information. In a Gaussian system this can be achieved based on combining the squared of the variances. The amount of uncertainty contained within the aggregated value is strongly determined by how correlated (or covarying) the uncertainties of the aggregated values are. However, we typically do not know what the correlations are in space or time. Due to this lack of information, we conservatively assume that the correlation is full (or completely correlated). We recognise that this assumption leads to an overestimation of the uncertainty reported in our aggregate numbers, but we consider this to be more reasonable and transparent than assuming an unfounded correlation length scale or to assume they are fully uncorrelated. Assuming no correlation (or that uncertainties are fully independent) leads to little or no accumulation of uncertainty as we aggregate in space or time. We have used our assumption, that uncertainties are fully correlated, in many peer-reviewed publications (e.g. Exbrayat et al., 2018b; Smallman et al., 2021; Williams et al., 2025).

Standard uncertainty approaches assume a Gaussian distribution which is inconsistent with our ensemble-based approach which often results in non-Gaussian uncertainties. Directly combining the pixel-level ensembles between pixels is equivalent to assuming fully independent uncertainty – which is the reverse of what we currently do as it leads to a radically overconfident estimate. Combining between pixels based on their quantiles is the equivalent of assuming fully correlated uncertainty.

Our key question focuses on the uncertainty at pixel scale for which the uncertainties are known explicitly from our ensemble-based approach and not impacted by the spatial aggregation question. Thus, a sensitivity analysis of the spatial correlations of uncertainty is out of scope for this study.

We will create a new section 2.6 which expands the fundamental basis of uncertainty, its reporting and propagation assumptions within the scope of this paper.

"Sec. 2.6 Approach to uncertainty

The CARDAMOM analysis generates ensembles of retrieved parameter for each pixel. The ensembles are parameters which are consistent with the assimilated information for that pixel. We run a subsample of the accepted parameters (300) through DALEC at each pixel to generate a corresponding ensemble of carbon cycle states and fluxes. It is these pixel-level ensembles from which we directly quantify the pixel-level uncertainty. We determine the quantiles for the 95 % and 68 % (equivalent to the standard deviation) confidence intervals as standard metrics. We use a quantile-based approach as the

MCMC algorithm we use does not assume a Gaussian distribution in its retrieved posteriors.

Our focus in on the pixel-level uncertainty. However, we provide global and regional level aggregated budgets to support comparisons with other studies. Aggregation requires propagation of the pixel-level uncertainties into the spatially aggregated estimates. However, we do not have adequate information to determine the spatial correlation of uncertainties. Thus, we assume that uncertainties are fully correlated in space giving the most conservative interpretation of uncertainties. We recognise that this will result in an overestimate of our uncertainty, but that this is preferable from presenting a false confidence."

Since data assimilation uses certain parameterizations of physical processes such as surface fluxes of mass/energy and plant phenology, the systematic errors of models and the input data may cancel through tuning and/or parameter calibration, leading to improved estimates of C stocks and fluxes. It is unclear what, e.g. more realistic physical models or quality of input data, are responsible for the improvement of the new global C stocks and fluxes data.

We recognise the reviewer's statement that model calibration (parameterisation) can achieve its fit to data through compensating errors, more broadly reflecting the challenges posed by equifinality. The reviewer also asks about whether the process representation, input data or other is leading to an improvement in the estimates of carbon stocks.

The equifinality and compensation error challenge is inherent in all models, whether process-based, statistical or machine learning methods. We deploy a range of approaches to mitigate against this challenge. Our ecological and dynamical constraints (Sec. 2.1.1) explicitly target unrealistic compensating errors and therefore reduce the scope of equifinality (Line 102) by rejecting parameter combinations which are unrealistic or generate unrealistic model behaviours. Second, we perform an evaluation using fully independent datasets (Sec. 2.6) on carbon emission from fire and net biome exchange to assess the quality of our calibration (Sec. 3.1, lines 248-253, figures 2 & 3).

On the final point about what is responsible for the improvement in the carbon stocks and flux estimates, we believe is a misunderstanding either in our interpretation of the comment or in the reviewer's read of the manuscript. We generate a single analysis, therefore, there is no comparison made between an existing analysis of the carbon cycle against which improvement is made.

The production of the monthly new C data was driven by monthly meteorological and disturbance information. This is problematic due to the nonlinear response of C-dynamics to environmental drivers.

DALEC is designed and constructed from the ground up to operate at daily to monthly time scales. As part of DALEC's development key ecological processes e.g. photosynthesis, respiration and evaporation have been calibrated using daily and longer time step information but evaluated using sub-daily in-situ estimates aggregated to the daily (e.g. Williams et al., 1997; Smallman & Williams 2019). These evaluations tell us that DALEC can realistically simulatie the aggregate daily fluxes.

Disturbance events will likely happen on time scales less than a monthly time step. However, we are not aware of any long-term global analyses for tree loss or fire which operates at less than annual or monthly time steps respectively. Thus, we are utilising the current state of the art which is appropriate for our stated research question.

A more severe limitation of this study is that the main objective of producing new C stocks and fluxes data is to reduce their uncertainty for the determination of global carbon sink/source distribution, among other purposes.

Unfortunately, the reviewer has misunderstood. Our objective is not to reduce uncertainty but to better quantify how much uncertainty there is to address our central research question, Specifically, *what* can contemporary observations tell us about terrestrial carbon cycling. This objective is stated in line 6-7, line 15-16, line 76-79 and line 343-344.

Furthermore, our explicitly stated outputs of the study are on reporting budgets, diagnosing the determinants of net exchange, characterising sources of uncertainty (not reducing) and determining C allocation and residence times. See below for full quotes:

"1) Report global and regional C budgets aligned to the RECCAP2 regions.
2) Diagnose the main determinants of net C exchange at global and regional scales aligned to the RECCAP2 regions.
3) Characterise sources of uncertainty from across ecosystem components to the overall net exchange.
4) Determine allocation of photosynthate to plant tissues, and their respective C residence times."

Yet, focusing on uncertainty alone will not be sufficient to achieve the goal. Reducing biases is more important than reducing uncertainties. Imagine the estimates of wood and soil C storage have no uncertainty but substantial biases (systematic EGUSPHERE-

2026-3014 2 errors), which could lead to totally wrong sign and magnitude of net C sink/source. As a result, this effort missed the target.

We recognise that bias is a significant challenge and one we note in our introduction but refer to (i) as inconsistencies between datasets (line 34-36) and (ii) as large differences between different data products that attempt to map the same variable, explicitly highlighting the current generation of GPP products as an example (line 43-46). Thus, bias is not easily quantified or removed. We argue that efforts such as ours are essential to understand what systemic representation of terrestrial carbon cycling is implied by these different datasets. Our methodology also lends itself to minimising the impacts of bias. By assimilating orthogonal datasets – each of which is likely biased as well as uncertain - into a systemic representation the resultant Bayesian analysis must balance these competing information streams. Therefore, the combined uncertainty is directly embedded into our uncertainty bounded analyses.

However, we also note that our analysis demonstrates that the actual magnitude of e.g. wood stock is not strongly correlated with the uncertainty in net exchange. Rather it is the dynamics implied by these datasets which is most informative (Fig. 7) as is one of our stated conclusions.

The authors did not give a definite and clear answer to the question in the paper title. The Conclusions section only summarizes the work and discusses briefly the potential applications of the new C data.

We are sorry that the reviewer missed this but we explicitly answer the overall question in the abstract (line 15), the discussion line 353 "Our results show that current global datasets for terrestrial ecosystems are not adequate to understand global C dynamics - in particular a critical lack of knowledge on the dynamics of soils*"* and again in the conclusions line 459 "Existing state-of-the-art geospatial information on LAI, fAPAR, woody biomass, soil C and LCA are inadequate to confidently (>95% CI) estimate net C sources and sinks across much of the vegetated land surface".

This paper reads more like a technical report than a research article. If this paper is meant to be a technical report, using a question as the title may not be appropriate.

Respectfully we disagree. The manuscript addresses a key question on what ecological information is contained is the current state-of-the-art EO datasets which are being used to support land management decision making and the development of national and international policy. Furthermore, we identify and prioritise key knowledge gaps and hypotheses on datasets to address these gaps.

References

Dietze, M.C. (2017), Prediction in ecology: a first-principles framework. Ecol Appl, 27: 2048-2060. https://doi.org/10.1002/eap.1589

van de Schoot, R., Depaoli, S., King, R. *et al*. Bayesian statistics and modelling. *Nat Rev Methods Primers* **1**, 1 (2021). https://doi.org/10.1038/s43586-020-00001-2

Smallman, T. L., , and Williams, M.: Description and validation of an intermediate complexity model for ecosystem photosynthesis and evapotranspiration: ACM-GPP-ETv1, Geosci. Model Dev., 12, https://doi.org/10.5194/gmd-12-2227-2019, 2019.

Stoy et al., (2006) An evaluation of models for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respiration, Agricultural and Forest Meteorology, https://doi.org/10.1016/j.agrformet.2006.09.001.

Williams, M., Rastetter, E., Fernandes, D., Goulden, M., Shaver, G., and Johnson, L.: Predicting gross primary productivity in terrestrial ecosystems, Ecol. Appl., 7, 882–894, 1997. a, b, c, d, e, f, g